SOURCE AND ACCURACY STATEMENT

for the 1993 Public Use Files
from the Survey of Income and Program Participation (SIPP)

SOURCE OF DATA

The SIPP universe is the noninstitutionalized resident population living in the United States. The population includes persons living in group quarters, such as dormitories, rooming houses, and religious group dwellings. Not eligible to be in the survey are crew members of merchant vessels, Armed Forces personnel living in military barracks, and institutionalized persons, such as correctional facility inmates and nursing home residents. Also, not eligible are United States citizens residing abroad. Foreign visitors who work or attend school in this country and their families are eligible; all others are not eligible. With the exceptions noted above, field representatives interview eligible persons who are at least 15 years of age at the time of the interview.

The 1993 panel of the SIPP sample is located in 284 Primary Sampling Units (PSUs) each consisting of a county or a group of contiguous counties. Within these PSUs, we systematically selected expected clusters of two living quarters (LQs) from lists of addresses prepared for the 1980 decennial census to form the bulk of the sample. To account for LQs built within each of the sample areas after the 1980 census we selected a sample containing clusters of four LQs from permits issued for construction of residential LQs up until shortly before the beginning of the panel.

In jurisdictions that have incomplete addresses or don't issue building permits, we sampled small land areas, listed expected clusters of four LQs, and then subsampled. In addition, we selected a sample of LQs from a supplemental frame that included LQs identified as missed in the 1980 census.

Approximately 27,300 living quarters were originally designated for the 1993 panel. For Wave 1 of the panel, we obtained interviews from occupants of about 19,900 of the 27,300 designated living quarters. We found most of the remaining 7,400 living quarters in the panel to be vacant, demolished, converted to nonresidential use, or otherwise ineligible for the survey. However, we did not interview approximately 2,000 of the 7,400 living quarters in the panel because the occupants refused to be interviewed, could not be found at home, were temporarily absent, or were otherwise unavailable. Thus, occupants of about 91 percent of all eligible living quarters participated in the first interview of the panel.

For subsequent interviews, only original sample persons (those in Wave 1 sample households and interviewed in Wave 1) and persons living with them are eligible to be interviewed. We followed original sample persons if they moved to a new address, unless the new address was more than 100 miles from a SIPP sample area, we attempted telephone interviews. When original sample persons moved to remote parts of the country and were unreachable by

telephone, moved without leaving a forwarding address, or refused the interview, additional noninterviews resulted.

The Bureau divides sample households within a given panel into four subsamples of nearly equal size. We call these subsamples rotation groups 1, 2, 3, or 4 and interview one rotation group each month. Beginning in February 1993, we schedule interviews for each household in the sample at 4 month intervals over a period of roughly 2½ years. The reference period for the questions is the 4-month period preceding the interview month. A wave is one cycle of four interviews covering the entire sample, using the same questionnaire.

A unique feature of the SIPP design is overlapping panels. The overlapping design allows combining of panels and essentially doubles the sample size. It is possible to combine selected interviews for the 1993 panels with interviews from the 1992 panels. We include information necessary to do this later in this statement.

The public use files include core and supplemental (topical module) data. Field representatives repeat core questions at each interview over the life of the panel. Topical modules include questions which are asked only in certain waves. The 1993 and 1992 panel topical modules are shown in tables 1 and 2 respectively.

Tables 3 and 4 indicate the reference months and interview months for the collection of data from each rotation group for the 1993 and 1992 panels respectively. For example, Wave 1 rotation group 2 of the 1993 panel was interviewed in February 1993 and data for the reference months October 1992 through January 1993 were collected.

Estimation. We derived SIPP person weights in each panel from several stages of weight adjustments. In the first wave, we gave each person a base weight equal to the inverse of his/her probability of selection. For each subsequent interview, the Bureau gave each person a base weight that accounted for following movers.

We applied a factor to each interviewed person's weight to account for the SIPP sample areas not having the same population distribution as the strata they are from.

We applied a noninterview adjustment factor to the weight of every occupant of interviewed households to account for persons in noninterviewed occupied households which were eligible for the sample. (The Bureau treated individual nonresponse within partially interviewed households with imputation. We made no special adjustment for noninterviews in group quarters.)

The Bureau used complex techniques to adjust the weights for nonresponse. For a further explanation of the techniques used, see the Nonresponse Adjustment Methods for Demographic Surveys at the U.S. Bureau of the Census, November 1988, Working paper 8823, by R. Singh and R. Petroni. The success of these techniques in avoiding bias is unknown. An example of successfully avoiding bias can be found in "Current Nonresponse Research for the Survey of Income and Program Participation" (paper by Petroni, presented at the Second International Workshop on Household Survey Nonresponse, October 1991).

We performed an additional stage of adjustment to persons' weights to reduce the mean square errors of the survey estimates. We accomplished this by ratio adjusting the sample estimates to agree with monthly Current Population Survey (CPS) type estimates of the civilian (and some military) noninstitutional population of the United States at the national level by demographic characteristics including age, sex, and race as of the specified date. The Bureau brought CPS estimates by age, sex, and race into agreement with adjusted estimates from the 1990 decennial census. Adjustments to the 1990 decennial census estimates include an adjustment for undercount[1] and also reflect births, deaths, immigration, emigration, and changes in the Armed Forces since 1990. The 1991 panel wave 6 is the first panel and wave to use the 1990 census based controls in the weighting. Weights for earlier waves were based on independent population estimates derived by updating the 1980 decennial census counts. For information about the effect of the new population controls on various person and household characteristics, refer to tables 5 through 10 from the January 10, 1994 memorandum for Turner from Waite, titled "SIPP 91: Source and Accuracy Statement for 1991 Wave 6+ Panel Public Use Files." In addition, we controlled SIPP estimates to independent Hispanic controls and made an adjustment to assign equal weights to husbands and wives within the same household. We implemented all of the above adjustments for each reference month and the interview month.

Use of Weights. Each household and each person within each household on each wave tape has five weights. Four of these weights are reference month specific and therefore can be used only to form reference month estimates. Average reference month estimates to form estimates of monthly averages over some period of time. For example, using the proper weights, one can estimate the monthly average number of households in a specified income range over November and December 1993. To estimate monthly averages of a given measure (e.g., total, mean) over a number of consecutive months, sum the monthly estimates and divide by the number of months.

The remaining weight is interview month specific. Use this weight to form estimates that specifically refer to the interview month (e.g., total persons currently looking for work), as well as estimates referring to the time period including the interview month and all previous months (e.g., total persons who have ever served in the military).

To form an estimate for a particular month, use the reference month weight for the month of interest, summing over all persons or households with the characteristic of interest whose reference period includes the month of interest. Multiply the sum by a factor to account for the number of rotations contributing data for the month. This factor equals four divided by the number of rotations contributing data for the month. For example, December 1992 data is only available from rotations 2, 3, and 4 for Wave 1 of the 1993 panel (see table 3), so apply a factor of 4/3. To form an estimate for an interview month, use the procedure discussed above using the interview month weight provided on the file.

---

[1]   See "The 1990 Post-Enumeration Survey: Operations and Results" by Howard Hogan in the 1993 Proceedings of the Undercount in the 1990 Census Section, American Statistical Association.

Apply factors greater than 1 when constructing estimates for months with four rotations worth of data from a wave file. However, when using core data from consecutive waves together, data from all four rotations may be available, in which case the factors are equal to 1.

These tapes contain no weight for characteristics that involve a persons's or household's status over two or more months (e.g., number of households with a 50 percent increase in income between November and December 1992).

Producing Estimates for Census Regions and States. The total estimate for a region is the sum of the state estimates in that region. Using this sample, estimates for individual states are subject to very high variance and are not recommended. The state codes on the file are primarily of use for linking respondent characteristics with appropriate contextual variables (e.g., state-specific welfare criteria) and for tabulating data by user-defined groupings of states.

Producing Estimates for the Metropolitan Population. For Washington, DC and 18 states, we identify metropolitan or non-metropolitan residence (variable H*-METRO). In 28 additional states, where the non-metropolitan population in the sample was small enough to present a disclosure risk, we recoded a fraction of the metropolitan sample to be indistinguishable from non-metropolitan cases (H*-METRO=2). In these states, therefore, the cases coded as metropolitan (H*-METRO=1) represent only a subsample of that population.

In producing state estimates for a metropolitan characteristic, multiply the individual, family, or household weights by the metropolitan inflation factor for that state, presented in table 5. (This inflation factor compensates for the subsampling of the metropolitan population and is 1.0 for the states with complete identification of the metropolitan population.)

The same procedure applies when creating estimates for particular identified MSA's or CMSA's--apply the factor appropriate to the state. For multi-state MSA's, use the factor appropriate to each state part. For example, to tabulate data for the Washington, DC-MD-VA MSA, apply the Virginia factor of 1.0321 to weights for residents of the Virginia part of the MSA; Maryland and DC residents require no modification to the weights (i.e., their factors equal 1.0).

In producing regional or national estimates of the metropolitan population, it is also necessary to compensate for the fact that we don't identify a metropolitan subsample within one state (West Virginia). Thus, use factors in the right-hand column of table 11 for regional and national estimates. The results of regional and national tabulations of the metropolitan population will be biased slightly. However, less than one-half of one percent of the metropolitan population is not represented.

Producing Estimates for the Non-Metropolitan Population. State, regional, and national estimates of the non-metropolitan population cannot be computed directly, except for Washington, DC and the 18 states where the factor for state tabulations in table 5 is 1.0. In all other states, the cases identified as not in the metropolitan subsample (METRO=2) are a mixture of non-metropolitan and metropolitan households. Only an indirect method of

estimation is available: first compute an estimate for the total population, then subtract the estimates for the metropolitan population. The results of these tabulations will be slightly biased.

Combined Panel Estimates. Both the 1993 and 1992 panels provide data for October 1992-April 1995. Thus, obtain estimates for these time periods by combining the corresponding panels. However, since the Wave 1 questionnaire differs from the subsequent waves' questionnaire and since the procedures changed between the 1992 and 1993 panels, we recommend that estimates not be obtained by combining Wave 1 data of the 1993 panel with data from another panel. In this case, use the estimate obtained from either panel. Additionally, even for other waves, care should be taken when combining data from two panels since questionnaires for the two panels differ somewhat and since the length of time in sample for interviews from the two panels differ.

Obtain combined panel estimates either (1) by combining estimates derived separately for the two panels or (2) by first combining data from the two files and then producing an estimate.

1.   Combining Separate Estimates

Combine corresponding estimates from two consecutive year panels to create joint estimates by using the formula

$$\hat{J} = W\hat{J}_1 + (1-W)\hat{J}_2 \tag{A}$$

$\hat{J}$ = joint estimate (total, mean, proportion, etc);

$\hat{J}_1$ = estimate from the earlier panel;

$\hat{J}_2$ = estimate from the later panel;

W = weighting factor of the earlier panel.

To combine the 1992 and 1993 panels use a W value of 0.517 unless one of the panels contributes no information to the estimate. In that case, assign the panel contributing information a factor of 1. Assign the other a factor of zero.

2.   Combining Data from Separate Files

Start by first creating a file containing the data from the two panel files. Apply the weighting factor, W, to the weight of each person from the earlier panel and apply (1-W) to the weight of each person from the later panel. Then produce estimates using the same methodology as used to obtain estimates from a single panel.

Illustration for computing combined panel estimate.

Suppose SIPP estimates for Wave 5, 1992 panel show there were 441,000 households with monthly May income above $6,000. Also, suppose SIPP estimates for Wave 2, 1993 panel show there were 435,000 households with monthly May income above $6,000. Using formula (A), the joint level estimate is

$$\hat{J} = (0.517)(441,000) + (0.483)(435,000) = 438,000$$

ACCURACY OF ESTIMATES

We base SIPP estimates on a sample. The sample estimates may differ somewhat from the values obtained from administering a complete census using the same questionnaire, instructions, and enumerators. The difference occurs because with an estimate based on a sample survey two types of errors are possible: nonsampling and sampling. We can provide estimates of the magnitude of the SIPP sampling error, but this is not true of nonsampling error. The next few sections describe SIPP nonsampling error sources, followed by a discussion of sampling error, its estimation, and its use in data analysis.

Nonsampling Variability. We attribute nonsampling errors to many sources, they include:

- inability to obtain information about all cases in the sample,
- definitional difficulties,
- differences in the interpretation of questions,
- inability or unwillingness on the part of the respondents to provide correct information,
- inability to recall information,
- errors made in collection (e.g. recording or coding the data),
- errors made in processing the data,
- errors made in estimating values for missing data,
- biases resulting from the differing recall periods caused by the interviewing pattern used,
- undercoverage.

We used quality control and edit procedures to reduce errors made by respondents, coders and interviewers. More detailed discussions of the existence and control of nonsampling errors in the SIPP are in the SIPP Quality Profile.

Undercoverage in SIPP resulted from missed living quarters and missed persons within sample households. It is known that undercoverage varies with age, race, and sex. Generally, undercoverage is larger for males than for females and larger for Blacks than for Nonblacks. Ratio estimation to independent age-race-sex population controls partially corrects for the bias due to survey undercoverage. However, biases exist in the estimates when persons in missed households or missed persons in interviewed households have characteristics different from those of interviewed persons in the same age-race-sex group.

A common measure of survey coverage is the coverage ratio, the estimated population before ratio adjustment divided by the independent population control. Table 6 shows CPS coverage ratios for age-sex-race groups for 1992. The CPS coverage ratios can exhibit some variability from month to month, but these are a typical set of coverage ratios. Other Census Bureau household surveys like the SIPP experience similar coverage.

Comparability with Other Estimates. Exercise caution when comparing data from this report with data from other SIPP publications or with data from other surveys. Comparability problems are from varying seasonal patterns for many characteristics, different nonsampling errors, and different concepts and procedures. Refer to the SIPP Quality Profile for known differences with data from other sources and further discussion.

Sampling Variability. Standard errors indicate the magnitude of the sampling error. They also partially measure the effect of some nonsampling errors in response and enumeration, but do not measure any systematic biases in the data. The standard errors mostly measure the variations that occurred by chance because we surveyed a sample rather than the entire population.

USES AND COMPUTATION OF STANDARD ERRORS

Confidence Intervals. The sample estimate and its standard error enable one to construct confidence intervals, ranges that would include the average result of all possible samples with a known probability. For example, if we selected all possible samples and surveyed each of these under essentially the same conditions and with the same sample design, and if we calculated an estimate and its standard error from each sample, then:

1. Approximately 68 percent of the intervals from one standard error below the estimate to one standard error above the estimate would include the average result of all possible samples.

2. Approximately 90 percent of the intervals from 1.645 standard errors below the estimate to 1.645 standard errors above the estimate would include the average result of all possible samples.

3. Approximately 95 percent of the intervals from 1.960 standard errors below the estimate to 1.960 standard errors above the estimate would include the average result of all possible samples.

The average estimate derived from all possible samples is or is not contained in any particular computed interval. However, for a particular sample, one can say with a specified confidence that the confidence interval includes the average estimate derived from all possible samples.

Hypothesis Testing. One may also use standard errors for hypothesis testing. Hypothesis testing is a procedure for distinguishing between population characteristics using sample estimates. The most common type of hypothesis tested is 1) the population characteristics are identical versus 2) they are different. One can perform tests at various levels of significance, where a level of significance is the probability of concluding that the characteristics are different when, in fact, they are identical.

Unless noted otherwise, all statements of comparison in the report passed a hypothesis test at the 0.10 level of significance or better. This means that, for differences cited in the report, the estimated absolute difference between parameters is greater than 1.645 times the standard error of the difference.

To perform the most common test, compute the difference $X_A - X_B$, where X_A and X_B are sample estimates of the characteristics of interest. A later section explains how to derive an estimate of the standard error of the difference $X_A - X_B$. Let that standard error be s_{DIFF}. If $X_A - X_B$ is between -1.645 times s_{DIFF} and +1.645 times s_{DIFF}, no conclusion about the characteristics is justified at the 10 percent significance level. If, on the other hand, $X_A - X_B$ is smaller than -1.645 times s_{DIFF} or larger than +1.645 times s_{DIFF}, the observed difference is significant at the 10 percent level. In this event, it is commonly accepted practice to say that the characteristics are different. Of course, sometimes this conclusion will be wrong. When the characteristics are, in fact, the same, there is a 10 percent chance of concluding that they are different.

Note that as we perform more tests, more erroneous significant differences will occur. For example, at the 10 percent significance level, if we perform 100 independent hypothesis tests in which there are no real differences, it is likely that about 10 erroneous differences will occur. Therefore, interpret the significance of any single test cautiously.

Note Concerning Small Estimates and Small Differences. We show summary measures in the report only when the base is 200,000 or greater. Because of the large standard errors involved, there is little chance that estimates will reveal useful information when computed on a base smaller than 200,000. Also, nonsampling error in one or more of the small number of cases providing the estimate can cause large relative error in that particular estimate. We show estimated numbers, however, even though the relative standard errors of these numbers are larger than those for the corresponding percentages. We provide smaller estimates primarily to permit such combinations of the categories as serve each user's needs. Therefore, be careful in the interpretation of small differences since even a small amount of nonsampling error can cause a borderline difference to appear significant or not, thus distorting a seemingly valid hypothesis test.

Standard Error Parameters and Tables and Their Use. Most SIPP estimates have greater standard errors than those obtained through a simple random sample because we sampled clusters of living quarters for the SIPP. To derive standard errors at a moderate cost and applicable to a wide variety of estimates, we made a number of approximations. We grouped estimates with similar standard error behavior and developed two parameters (denoted "a" and "b") to approximate the standard error behavior of each group of estimates. Because the actual standard error behavior was not identical for all estimates within a group, the standard errors we computed from these parameters provide an indication of the order of magnitude of the standard error for any specific estimate. These "a" and "b" parameters vary by characteristic and by demographic subgroup to which the estimate applies. Use base "a" and "b" parameters found in table 7 for 1993 panel estimates. Note that for estimates which include data for wave 5 and beyond multiply the "a" and "b" parameters by 1.09 to account for sample attrition.

The factors provided in table 8 when multiplied by the base parameters of table 7 for a given subgroup and type of estimate give the "a" and "b" parameters for that subgroup and estimate type for the specified reference period. For example, the base "a" and "b" parameters for total number of households are -0.0000702 and 6,715, respectively. For Wave 1 the factor for October 1992 is 4 since only 1 rotation month of data is available. So, the "a" and "b" parameters for total household income in October 1992 based on Wave 1 are -0.0002808 and 26,860, respectively. Also for Wave 1, the factor for the first quarter of 1993 is 1.2222 since 9 rotation months of data are available (rotations 1 and 4 provide 3 rotations months each, while rotations 2 and 3 provide 1 and 2 rotation months, respectively). So the "a" and "b" parameters for total number of households in the first quarter of 1993 are -0.0000857 and 8,207, respectively for Wave 1.

Use the "a" and "b" parameters to calculate the standard error for estimated numbers and percentages. Because the actual standard error behavior was not identical for all estimates within a group, the standard errors computed from these parameters provide an indication of the order of magnitude of the standard error for any specific estimate. The following sections give methods for using these parameter for computation of approximate standard errors.

For users who wish further simplification, we also provide general standard errors in tables 9 through 12. Note that these standard errors only apply when data from all four rotations are used and you need to adjust these standard errors by a factor from table 7. The standard errors resulting from this simplified approach are less accurate. Methods for using these parameters and tables for computation of standard errors are given in the following sections.

For the 1992, 1993 combined panel parameters, multiply the parameters in table 7 by the appropriate factor from table 16.   The factors provided in table 17 adjust parameters for the number of rotation months available for a given estimate. These factors, when multiplied by the combined panel parameters derived from table 7 for a given subgroup and type of estimate, give the "a" and "b" parameters for that subgroup and estimate type for the specified combined reference period.

Table 13 provides base "a" and "b" parameters for calculating 1993 topical module variances. Table 14 provides base "a" and "b" parameters for computing the 1992, 1993 combined panel topical module variances.

Described below are procedures for calculating standard errors for the types of estimates most commonly used. Note specifically that these procedures apply only to reference month estimates or averages of reference month estimates. Refer to the section "Use of Weights" for a more detailed discussion of the construction of estimates. We included stratum codes and half sample codes on the tapes so users can compute variances directly by methods such as balanced repeated replications (BRR). William G. Cochran provides a list of references discussing the application of this technique. (See Sampling Techniques, 3rd Ed., New York: John Wiley and Sons, 1977, p. 321.)

Standard errors of estimated numbers. Obtain the approximate standard error, s_x, of an estimated number of persons, households, families, unrelated individuals and so forth, in one of two ways. Both apply when data from all four rotations are used to make the estimate. However, only the second method should be used when less than four rotations of data are available for the estimate. Note that neither method should be applied to dollar values.

The standard error may be obtained by the use of the formula

$$s_x = fs \tag{1}$$

where f is the appropriate "f" factor from table 7, and s is the standard error on the estimate obtained by interpolation from table 9 or 10. Alternatively, approximate s_x using the formula,

$$s_x = \sqrt{ax^2 + bx} \tag{2}$$

from which we calculated the standard errors in tables 9 and 10. Here x is the size of the estimate and "a" and "b" are the parameters associated with the particular type of characteristic. Use of formula 2 will provide more accurate results than the use of formula 1.

Illustration.

Suppose SIPP estimates for Wave 1 of the 1993 panel show that there were 472,000 black households with monthly household income above $6,000. The appropriate parameters and factor from table 7 and the appropriate general standard error from table 9 are

   a = -0.0004187   b = 4,640   f = 0.83   s = 55,000

Using formula 1, the approximate standard error is

$s_x = 46,000$

Using formula 2, the approximate standard error is

$$\sqrt{(-0.0004187)(472,000)^2 + (4,640)(472,000)} = 46,000$$

Using the standard error based on formula 2, the approximate 90-percent confidence interval as shown by the data is from 396,000 to 548,000. Therefore, a conclusion that the average estimate derived from all possible samples lies within a range computed in this way would be correct for roughly 90% of all samples.

Illustration for computing standard errors for combined panel estimates.

Suppose the combined SIPP estimate for total number of males in the 16+ Income and Labor Force for Wave 6, 1992 panel and Wave 3, 1993 panel was 92,398,000. The combined panel parameters for total males are obtained by multiplying the appropriate "a" and "b" values from table 7 by the appropriate factors from tables 16 and 17. The 1993 parameters and factors are a = -0.0000580, b = 5,433, g = 1.0000 and factor = 1.0000, respectively. Thus, the combined panel parameters are a = -0.0000580 and b = 5,433. Using formula 2, the approximate standard error is

$$S = \sqrt{(-0.0000580)(92,398,000)^2 + (5,433)(92,398,000)} = 83,000$$

Standard Error of a Mean. Define a mean as the average quantity of some item (other than persons, families, or households) per person, family or household. For example, it could be the average monthly household income of females age 25 to 34. Use formulas below to approximate the standard error of a mean. Because of the approximations used in developing formula 3, an estimate of the standard error of the mean obtained from this formula will generally underestimate the true standard error. The formula used to estimate the standard error of a mean $\overline{x}$ is

$$s_{\overline{x}} = \sqrt{\left(\frac{b}{y}\right)s^2} \tag{3}$$

where y is the size of the base, s^2 is the estimated population variance of the item and b is the parameter associated with the particular type of item.

Estimate the population variance s^2 by one of two methods. In both methods we assume x_i is the value of the item for unit i. (Unit may be person, family, or household). To use the first method, divide the range of values for the item into c intervals. The upper and lower boundaries of interval j are Z_{j-1} and Z_j, respectively. Place each unit into one of c groups such that $Z_{j-1} < x_i \leq Z_j$.

The estimated population variance, s^2, is given by the formula:

$$s^2 = \sum_{j=1}^{c} p_j m_j^2 - \overline{x}^2,$$ (4)

where p_j is the estimated proportion of units in group j, and $m_j = (Z_{j-1} + Z_j)/2$. We assume the most representative value of the item in group j is m_j. If group c is open-ended, i.e., no upper interval boundary exists, then an approximate value for m_c is

$$m_c = \frac{3}{2} Z_{c-1}.$$

Compute the mean, $\overline{x}$, using the following formula:

$$\overline{x} = \sum_{j=1}^{c} p_j m_j.$$

In the second method, the estimated population variance is given by

$$s^2 = \frac{\sum_{i=1}^{n} w_i x_i^2}{\sum_{i=1}^{n} w_i} - \overline{x}^2,$$ (5)

where there are n units with the item of interest and w_i is the final weight for unit i.
Compute the mean, $\overline{x}$, using the formula

$$\overline{x} = \frac{\sum_{i=1}^{n} w_i x_i}{\sum_{i=1}^{n} w_i} \ .$$

When forming combined estimates using formula (A) from the section on combined panel estimates, calculate s^2, given by formula (4), by forming a distribution for each panel. Divide the range of values for the item into intervals. Obtain combined estimates for each interval using formula (A). Apply formula (4) to the combined distribution. To calculate $\overline{x}$ and s^2 given by formula (5), replace x_i by Wx_i for x_i from the earlier panel and $(1-W)x_i$ for x_i from the later panel.

Illustration.

Suppose that based on Wave 1 data, the distribution of monthly cash income for persons age 25 to 34 during the month of January 1993 is given in table 15.

Using formula 4 and the mean monthly cash income of $2,530 the approximate population

variance, s^2, is

$$s^2 = \left(\frac{1,371}{39,851}\right) (150)^2 + \left(\frac{1,651}{39,851}\right) (450)^2 + \ldots \ +$$

$$\left(\frac{1,493}{39,851}\right) (9,000)^2 - (2,530)^2 = 3,159,887 .$$

Using formula 3, the appropriate base "b" parameter and factor from table 7, the estimated standard error of a mean $\overline{x}$ is

$$s_{\overline{x}} = \sqrt{\left(\frac{5,433}{39,851,000}\right) (3,159,887)} = \$21$$

Standard error of an aggregate. We define an aggregate as the total quantity of an item summed over all the units in a group. Approximate the standard error of an aggregate using formula 6.

Because of the approximations used in developing formula (6), it will generally underestimate the true standard error. Let y be the size of the base, s^2 be the estimated population variance of the item obtained using formula (4) or (5) and b be the parameter associated with the particular type of item. The standard error of an aggregate is:

$$s_x = \sqrt{(b)\ (y)\ s^2} \tag{6}$$

Standard Errors of Estimated Percentages. The reliability of an estimated percentage, computed using sample data for both numerator and denominator, depends on the size of the percentage and its base. Estimated percentages are relatively more reliable than the corresponding estimates of the numerators of the percentages, particularly if the percentages are 50 percent or more, e.g., the percent of people employed is more reliable than the estimated number of people employed. When the numerator and denominator of the percentage have different parameters, use the parameter (and appropriate factor) of the numerator. If proportions are presented instead of percentages, note that the standard error of a proportion is equal to the standard error of the corresponding percentage divided by 100.

We commonly estimate two types of percentages. The first is the percentage of persons, families or households sharing a particular characteristic such as the percent of persons owning their own home. The second type is the percentage of money or some similar concept held by a particular group of persons or held in a particular form. Examples are the percent of total wealth held by persons with high income and the percent of total income received by persons on welfare.

For the percentage of persons, families, or households, calculate the approximate standard error, $s_{(x,p)}$, of an estimated percentage p using the formula

$$s_{(x,p)} = fs \tag{7}$$

when estimating p using data from all four rotations.

In this formula, f is the appropriate "f" factor from table 7 and s is the standard error of the estimate from table 11 or 12.

Alternatively, approximate it by the formula:

$$s_{(x,p)} = \sqrt{\frac{b}{x} \; (p) \; (100-p)}$$ (8)

from which we calculated the standard errors in tables 11 and 12. Here x is the size of the subclass of social units which is the base of the percentage, p is the percentage (0<p<100), and b is the parameter associated with the characteristic in the numerator. Using this formula gives more accurate results than using formula 7 above. Use this formula to estimate p for data with less than four rotations.

Illustration.

Suppose that, in the month of January 1993, 6.7 percent of the 16,812,000 persons in nonfarm households with a mean monthly household cash income of $4,000 to $4,999, were black. Using formula 8 and the "b" parameter of 7,310 from table 7 and a factor of 1 for the month of January 1993 from table 8, the approximate standard error is

$$\sqrt{\frac{7,310}{(16,812,000)} \; (6.7) \; (100-6.7)} = 0.52 \; percent$$

Consequently, the 90 percent confidence interval as shown by these data is from 5.8 to 7.6 percent.

Percentages of money require a more complicated formula. Estimate a percentage of money one of two ways. It may be the ratio of two aggregates:

$$p_I = 100 \; (X_A \; / \; X_N)$$

or it may be the ratio of two means with an adjustment for different bases:

$$p_I = 100 \; (\hat{P}_A \; \overline{X}_A \; / \; \overline{X}_N)$$

where x_A and x_N are aggregate money figures, $\overline{x}_A$ and $\overline{x}_N$ are mean money figures, and $\hat{p}_A$ is the estimated number in group A divided by the estimated number in group N. In either case, we estimate the standard error as

$$s_I = \sqrt{\left(\frac{\hat{p}_A \overline{x}_A}{\overline{x}_N}\right)^2 \left[\left(\frac{s_p}{\hat{p}_A}\right)^2 + \left(\frac{s_A}{\overline{x}_A}\right)^2 + \left(\frac{s_B}{\overline{x}_N}\right)^2\right]} \quad , \tag{9}$$

where s_p is the standard error of $\hat{p}_A$, s_A is the standard error of $\overline{x}_A$ and s_B is the standard error of $\overline{x}_N$. To calculate s_p, use formula 8. Calculate the standard errors of $\overline{x}_N$ and $\overline{x}_A$ using formula 3.

Note that there is frequently some correlation between $\hat{p}_A$, $\overline{x}_N$, and $\overline{x}_A$. Depending on the magnitude and sign of the correlations, the standard error will be over or underestimated.

Illustration.

Suppose that in January 1993, 9.8% of the households own rental property, the mean value of rental property is $72,121, the mean value of assets is $78,734, and the corresponding standard errors are 0.31%, $5799, and $2867. In total there are 86,790,000 households. Then, the percent of all household assets held in rental property is

$$= 100 \left((0.098)\frac{72121}{78734}\right) = 9.0\%$$

Using formula (9), the appropriate standard error is

$$s_I = \sqrt{\left(\frac{(0.098)(72121)}{78734}\right)^2 \left[\left(\frac{0.0031}{0.098}\right)^2 + \left(\frac{5799}{72121}\right)^2 + \left(\frac{2867}{78734}\right)^2\right]}$$

$$= 0.008$$

$$= 0.8\%$$

Standard Error of a Difference. The standard error of a difference between two sample estimates, x and y, is approximately equal to

$$S_{(x-y)} = \sqrt{S_x^2 + S_y^2} \qquad\qquad (10)$$

where s_x and s_y are the standard errors of the estimates x and y.

The estimates can be numbers, percents, ratios, etc. The above formula assumes that the correlation coefficient between the characteristics estimated by x and y is zero. If the correlation is really positive (negative), then this assumption will tend to cause overestimates (underestimates) of the true standard error.

Illustration.

Suppose that SIPP estimates show the number of persons age 35-44 years with monthly cash income of $4,000 to $4,999 was 3,186,000 in the month of January 1993 and the number of persons age 25-34 years with monthly cash income of $4,000 to $4,999 in the same time period was 2,619,000. Then, using parameters from table 7 and formula 2, the standard errors of these numbers are approximately 130,000 and 118,000, respectively. The difference in sample estimates is 567,000 and, using formula 10, the approximate standard error of the difference is

$$\sqrt{(130,000)^2 + (118,000)^2} = 176,000$$

Suppose that it is desired to test at the 10 percent significance level whether the number of persons with monthly cash income of $4,000 to $4,999 was different for persons age 35-44 years than for persons age 25-34 years. To perform the test, compare the difference of 567,000 to the product 1.645 x 176,000 = 290,000. Since the difference is greater than 1.645 times the standard error of the difference, the data show that the two age groups are significantly different at the 10 percent significance level.

Standard Error of a Median. The median quantity of some item such as income for a given group of persons, families, or households is that quantity such that at least half the group have as much or more and at least half the group have as much or less. The sampling variability of an estimated median depends upon the form of the distribution of the item as well as the size of the group. Use the procedure described below to calculate standard errors on medians.

An approximate method for measuring the reliability of an estimated median is to determine a confidence interval about it. (See the section on sampling variability for a general discussion of confidence intervals.) Use the following procedure to estimate the 68-percent confidence limits and hence the standard error of a median based on sample data.

1. Determine, using either formula 7 or formula 8, the standard error of an estimate of 50 percent of the group;

2. Add to and subtract from 50 percent the standard error determined in step 1;

3. Using the distribution of the item within the group, calculate the quantity of the item such that the percent of the group with more of the item is equal to the smaller percentage found in step 2. This quantity will be the upper limit for the 68-percent confidence interval. In a similar fashion, calculate the quantity of the item such that the percent of the group with more of the item is equal to the larger percentage found in step 2. This quantity will be the lower limit for the 68-percent confidence interval;

4. Divide the difference between the two quantities determined in step 3 by two to obtain the standard error of the median.

To perform step 3, you must interpolate. You may use different methods of interpolation. The most common are simple linear interpolation and Pareto interpolation. The appropriateness of the method depends on the form of the distribution around the median. If density is declining in the area, then we recommend Pareto interpolation. If density is fairly constant in the area, then we recommend linear interpolation. Never use Pareto interpolation if the interval contains zero or negative measures of the item of interest. Use interpolation as follows. The quantity of the item such that "p" percent have more of the item is

$$X_{pN} = \exp\left[\left(Ln\left(\frac{pN}{N_1}\right) \middle/ Ln\left(\frac{N_2}{N_1}\right)\right) Ln\left(\frac{A_2}{A_1}\right)\right]A_1 \qquad (11)$$

if Pareto Interpolation is indicated and

$$X_{pN} = \left[\frac{PN-N_1}{N_2-N_1} (A_2-A_1) + A_1\right] \qquad (12)$$

if linear interpolation is indicated, where

N	is the size of the group,
A_1 and A_2	are the lower and upper bounds, respectively, of the interval in which X_{pN} falls,
N_1 and N_2	are the estimated number of group members owning more than A_1 and A_2, respectively,
exp	refers to the exponential function and
Ln	refers to the natural logarithm function.

<u>Illustration.</u>

To illustrate the calculations for the sampling error on a median, we return to table 15. The median monthly income for this group is $2,158. The size of the group is 39,851,000.

1.   Using formula 8, the standard error of 50 percent on a base of 39,851,000 is about 0.6 percentage points.

2.   Following step 2, the two percentages of interest are 49.4 and 50.6.

3.   By examining table 15, we see that the percentage 49.4 falls in the income interval from 2000 to 2499. (Since 55.5% receive more than $2,000 per month, the dollar value corresponding to 49.4 must be between $2,000 and $2,500). Thus, $A_1 = \$2,000$, $A_2 = \$2,500$, $N_1 = 22,106,000$, and $N_2 = 16,307,000$.

In this case, we decided to use Pareto interpolation. Therefore, the upper bound of a 68% confidence interval for the median is

$$\$2,000 \exp\left[\left(Ln\left(\frac{(.494)(39,851,000)}{22,106,000}\right) \Big/ Ln\left(\frac{16,307,000}{22,106,000}\right)\right)Ln\left(\frac{2,500}{2,000}\right)\right] = \$2177$$

Also by examining table 15 , we see that 50.6 falls in the same income interval. Thus, A_1, A_2, N_1 and N_2 are the same. We also use Pareto interpolation for this case. So the lower bound of a 68% confidence interval for the median is

$$\$2,000 \exp\left[\left(Ln\left(\frac{(.506)(39,851,000)}{22,106,000}\right) \Big/ Ln\left(\frac{16,307,000}{22,106,000}\right)\right)Ln\left(\frac{2,500}{2,000}\right)\right] = \$2139$$

Thus, the 68-percent confidence interval on the estimated median is from $2139 to $2177. An approximate standard error is

$$\frac{\$2177 - \$2139}{2} = \$19$$

Standard Errors of Ratios of Means and Medians. Approximate the standard error for a ratio of means or medians by:

$$s_{\frac{x}{y}} = \sqrt{\left(\frac{x}{y}\right)^2 \left[\left(\frac{s_y}{y}\right)^2 + \left(\frac{s_x}{x}\right)^2\right]} \tag{13}$$

where x and y are the means or medians, and s_x and s_y are their associated standard errors. Formula 13 assumes that the means are not correlated. If the correlation between the population means estimated by x and y are actually positive (negative), then this procedure will tend to produce overestimates (underestimates) of the true standard error for the ratio of means.

Table 1. 1993 Panel Topical Modules

Wave	Topical Module
1	Recipiency History Employment History
2	Work Disability History Education and Training History Marital History Migration History Fertility History Household Relationships
3	Work Schedule Child Care Arrangements Child Support Agreements Support of Non-household Members Functional Limitations and Disability Utilization of Health Care Services
4	Selected Financial Assets Medical Expenses and Work Disability Real Estate, Shelter Costs, Dependent Care, and Vehicles
5	Taxes Annual Income and Retirement Accounts School Enrollment and Financing
6	To be decided in mid-1994
7	Eligibility Wealth
8	Taxes Annual Income and Retirement Accounts School Enrollment and Financing
9	To be decided in mid-1995
10	None

Table 2.   1992 Panel Topical Modules

Wave   Topical Module

1   Recipiency History
                Employment History

2   Work Disability, Education and Training, Marital, Migration History
                Fertility History
                Household Relationships

3   Extended Measures of Wellbeing
                    (Consumer Durables,
                    Living Conditions,
                    Basic Needs,
                    Expenditures,
                    Minimum Income)

4   Assets and Liabilities
                Retirement Expectations and Pension Plan Coverage
                Real Estate Property and Vehicles

5   Taxes
                Annual Income and Retirement Accounts
                School Enrollment and Financing

6   Child Care Arrangements
                Child Support Agreements
                Support of Non-household Members
                Functional Limitations and Disability
                Utilization of Health Care Services
                Work Schedule

7   Selected Financial Assets
                Medical Expenses and Work Disability
                Real Estate, Shelter Costs, Dependent Care,
                    and Vehicles

8   Taxes
                Annual Income and Retirement Accounts
                School Enrollment and Financing

9   To be decided in mid-1994

10   None

Table 3. Reference Months for Each Interview Month - 1993 Panel

Reference Period

Month of Interview	Wave/ Rotation	4th Quarter (1992) Oct Nov Dec			1st Quarter (1993) Jan Feb Mar			2nd Quarter (1993) Apr May Jun			3rd Quarter (1993) Jul Aug Sep			4th Quarter (1993) Oct Nov Dec			...	1st Quarter (1996) Jan Feb Mar			2nd Quarter (1996) Apr May Jun		
Feb 93	1/2	X	X	X	X																		
Mar	1/3		X	X	X	X																	
Apr	1/4			X	X	X	X																
May	1/1				X	X	X	X															
Jun	2/2					X	X	X	X														
Jul	2/3						X	X	X	X													
Aug	2/4							X	X	X	X												
Sept	2/1								X	X	X	X											
Oct	3/2									X	X	X	X										
Nov	3/3										X	X	X	X									
Dec	3/4											X	X	X	X								
.														. . .									
.														.			...						
.																	...	.					
May 96	10/1																		X	X	X	X		

Table 4. Reference Months for Each Interview Month - 1992 Panel

Reference Period

Month of Interview	Wave/ Rotation	4th Quarter (1991) Oct	Nov	Dec	1st Quarter (1992) Jan	Feb	Mar	2nd Quarter (1992) Apr	May	Jun	3rd Quarter (1992) Jul	Aug	Sep	4th Quarter (1992) Oct	Nov	Dec	...	1st Quarter (1995) Jan	Feb	Mar	2nd Quarter (1995) Apr	May	Jun
Feb 92	1/2	X	X	X	X																		
Mar	1/3		X	X	X	X																	
Apr	1/4			X	X	X	X																
May	1/1				X	X	X	X															
Jun	2/2					X	X	X	X														
Jul	2/3						X	X	X	X													
Aug	2/4							X	X	X	X												
Sept	2/1								X	X	X	X											
Oct	3/2									X	X	X	X										
Nov	3/3										X	X	X	X									
Dec	3/4											X	X	X	X								
.														.	.	.	...						
.															.		...						
.															.		...			.			
May 95	10/1																	X	X	X	X		

9—24

Table 5. **Metropolitan Subsample Factors to be Applied to Compute National and Subnational Estimates**

		Factors for use in State or CMSA (MSA) Tabulations	Factors for use in Regional or National Tabulations
Northeast:	Connecticut	1.0387	1.0387
	Maine	1.1609	1.1609
	Massachusetts	1.0000	1.0000
	New Hampshire	1.2234	1.2234
	New Jersey	1.0000	1.0000
	New York	1.0000	1.0000
	Pennsylvania	1.0000	1.0000
	Rhode Island	1.2506	1.2506
	Vermont	1.1609	1.1609
Midwest:	Illinois	1.0000	1.0000
	Indiana	1.0150	1.0150
	Iowa	1.1574	1.1574
	Kansas	1.2771	1.2771
	Michigan	1.0088	1.0088
	Minnesota	1.0366	1.0366
	Missouri	1.0364	1.0364
	Nebraska	1.3891	1.3891
	North Dakota	1.1574	1.1574
	Ohio	1.0000	1.0000
	South Dakota	1.1574	1.1574
	Wisconsin	1.0188	1.0188
South:	Alabama	1.1334	1.1389
	Arkansas	1.4784	1.4855
	Delaware	1.5593	1.5668
	D.C.	1.0000	1.0048
	Florida	1.0000	1.0048
	Georgia	1.0000	1.0048
	Kentucky	1.0168	1.0217
	Louisiana	1.0108	1.0157
	Maryland	1.0000	1.0048
	Mississippi	1.0000	1.0048
	North Carolina	1.0000	1.0048
	Oklahoma	1.0592	1.0643
	South Carolina	1.0073	1.0121
	Tennessee	1.0063	1.0112
	Texas	1.0064	1.0113
	Virginia	1.0321	1.0371
	West Virginia	- - -	- - -

- indicates no metropolitan subsample is identified for the state

Table 5 cont'd. Metropolitan Subsample Factors to be Applied to Compute National and Subnational Estimates

		Factors for use in State or CMSA (MSA) Tabulations	Factors for use in Regional or National Tabulations
West:	Alaska	1.4339	1.4339
	Arizona	1.0000	1.0000
	California	1.0000	1.0000
	Colorado	1.0571	1.0571
	Hawaii	1.0000	1.0000
	Idaho	1.4339	1.4339
	Montana	1.4339	1.4339
	Nevada	1.0000	1.0000
	New Mexico	1.0000	1.0000
	Oregon	1.0000	1.0000
	Utah	1.0000	1.0000
	Washington	1.0456	1.0456
	Wyoming	1.4339	1.4339

- indicates no metropolitan subsample is identified for the state

Table 6. 1992 CPS Coverage Ratios

Age	non-Black		Black		All Persons		
	Males	Females	Males	Females	Males	Females	Total
0-14	0.963	0.965	0.927	0.926	0.957	0.959	0.958
15	0.962	0.949	0.899	0.919	0.952	0.944	0.948
16	0.969	0.936	0.923	0.907	0.962	0.932	0.947
17	0.981	0.975	0.945	0.862	0.975	0.957	0.966
18	0.939	0.926	0.883	0.846	0.930	0.913	0.922
19	0.860	0.872	0.754	0.801	0.844	0.861	0.853
20-24	0.913	0.927	0.734	0.832	0.889	0.913	0.901
25-26	0.927	0.940	0.688	0.877	0.897	0.931	0.914
27-29	0.910	0.954	0.707	0.864	0.885	0.941	0.914
30-34	0.893	0.948	0.691	0.883	0.870	0.939	0.905
35-39	0.910	0.949	0.763	0.899	0.895	0.942	0.919
40-44	0.929	0.951	0.824	0.906	0.919	0.946	0.933
45-49	0.956	0.966	0.903	0.956	0.951	0.965	0.958
50-54	0.940	0.961	0.807	0.877	0.927	0.951	0.940
55-59	0.944	0.941	0.826	0.825	0.932	0.928	0.930
60-62	0.965	0.956	0.792	0.850	0.948	0.944	0.946
63-64	0.905	0.907	0.669	0.872	0.884	0.903	0.894
65-67	0.935	0.979	0.783	0.875	0.921	0.969	0.947
68-69	0.925	0.942	0.789	0.831	0.913	0.931	0.923
70-74	0.926	0.993	0.856	1.014	0.920	0.995	0.962
75-99	0.977	0.989	0.764	0.912	0.961	0.983	0.975
15+	0.928	0.953	0.782	0.883	0.912	0.944	0.929
0+	0.936	0.955	0.827	0.895	0.923	0.947	0.935

Table 7. SIPP Indirect Generalized Variance Parameters for the 1993 Panel

Characteristics[1]	Parameters		
PERSONS	**a**	**b**	**f**
Total or White			
16+ Program Participation(3)			
Both Sexes	-0.0000924	15,937	0.90
Male	-0.0001906	15,937	
Female	-0.0001794	15,937	
16+ Income and Labor Force (5)			
Both Sexes	-0.0000279	5,433	0.52
Male	-0.0000580	5,433	
Female	-0.0000537	5,433	
16+ Pension Plan[2] (4)			
Both Sexes	-0.0000510	9,950	0.71
Male	-0.0001061	9,950	
Female	-0.0000983	9,950	
All Others[2] (6)			
Both Sexes	-0.0000770	19,760	1.00
Male	-0.0001576	19,760	
Female	-0.0001504	19,760	
Black			
Poverty (1)			
Both Sexes	-0.0004182	13,594	0.83
Male	-0.0008952	13,594	
Female	-0.0007849	13,594	
All Others (2)			
Both Sexes	-0.0002249	7,310	0.61
Male	-0.0004814	7,310	
Female	-0.0004221	7,310	
HOUSEHOLDS			
Total or White	-0.0000702	6,715	1.00
Black	-0.0004187	4,640	0.83

[1] To account for sample attrition, multiply the a and b parameters by 1.09 for estimates which include data from Wave 5 and beyond.

For cross-tabulation, use the parameters of the characteristic with the smaller number within the parentheses.

[2] Use the "16+ Pension Plan" parameters for pension plan tabulations of persons 16+ in the labor force. Use the "All Others" parameters for retirement tabulations, 0+ program participation, 0+ benefits, 0+ income, and 0+ labor force tabulations, in addition to any other types of tabulations not specifically covered by another characteristic in this table.

Table 8.   Factors to be Applied to Table 7 Base Parameters to Obtain Parameters for Various Reference Periods

# of available rotation months[1]	factor
Monthly estimate	
1	4.0000
2	2.0000
3	1.3333
4	1.0000
Quarterly estimate	
6	1.8519
8	1.4074
9	1.2222
10	1.0494
11	1.0370
12	1.0000

---

[1]   The number of available rotation months for a given estimate is the sum of the number of rotations available for each month of the estimate.

Table 9. **Standard Errors of Estimated Numbers of Households, Families or Unrelated Persons** (Numbers in Thousands)

Size of Estimate	Standard Error[1]	Size of Estimate	Standard Error[1]
200	37	15,000	291
300	45	17,000	306
600	63	22,000	337
1,000	82	26,000	357
2,000	115	30,000	372
3,000	140	50,000	400
5,000	178	80,000	297
8,000	222	90,000	189
10,000	245	93,000	132
13,000	275	95,000	66

---

[1] To account for sample attrition, multiply the standard error of the estimate by 1.04 for estimates which include data from Wave 5 and beyond.

Table 10.   Standard Errors of Estimated Numbers of Persons (Numbers in Thousands)

Size of Estimate	Standard Error[1]	Size of Estimate	Standard Error[1]
200	63	26,000	679
300	77	30,000	724
600	109	50,000	892
1,000	140	80,000	1,043
2,000	198	100,000	1,098
3,000	242	130,000	1,126
5,000	311	150,000	1,110
8,000	391	180,000	1,031
10,000	436	200,000	934
13,000	494	230,000	687
15,000	528	240,000	554
17,000	560	250,000	357
22,000	630	256,000	111

[1]   To account for sample attrition, multiply the standard error of the estimate by 1.04 for estimates which include data from Wave 5 and beyond.

Table 11.   Standard Errors of Estimated Percentages of Households Families or Unrelated Persons

Base of Estimated Percentage (Thousands)	Estimated Percentages[1]					
	≤1 or ≥99	2 or 98	5 or 95	10 or 90	25 or 75	50
200	1.82	2.57	3.99	5.50	7.93	9.16
300	1.49	2.09	3.26	4.49	6.48	7.48
600	1.05	1.48	2.31	3.17	4.58	5.29
1000	0.82	1.15	1.79	2.46	3.55	4.10
2000	0.58	0.81	1.26	1.74	2.51	2.90
3000	0.47	0.66	1.03	1.42	2.05	2.37
5000	0.36	0.51	0.80	1.10	1.59	1.83
8000	0.29	0.41	0.63	0.87	1.25	1.45
10000	0.26	0.36	0.56	0.78	1.12	1.30
13000	0.23	0.32	0.50	0.68	0.98	1.14
15000	0.21	0.30	0.46	0.63	0.92	1.06
17000	0.20	0.28	0.43	0.60	0.86	0.99
22000	0.17	0.24	0.38	0.52	0.76	0.87
26000	0.16	0.22	0.35	0.48	0.70	0.80
30000	0.15	0.21	0.33	0.45	0.65	0.75
50000	0.12	0.16	0.25	0.35	0.50	0.58
80000	0.09	0.13	0.20	0.27	0.40	0.46
90000	0.09	0.12	0.19	0.26	0.37	0.43
93000	0.08	0.12	0.19	0.25	0.37	0.42
95000	0.08	0.12	0.18	0.25	0.36	0.42

---

[1]   To account for sample attrition, multiply the standard error of the estimate by 1.04 for estimates which include data from Wave 5 and beyond.

Table 12. Standard Errors of Estimated Percentages of Persons

Base of Estimated Percentage (Thousands)	Estimated Percentages[1]					
	≤1 or ≥99	2 or 98	5 or 95	10 or 90	25 or 75	50
200	3.13	4.40	6.85	9.43	13.61	15.72
300	2.55	3.59	5.59	7.70	11.11	12.83
600	1.81	2.54	3.96	5.44	7.86	9.07
1000	1.40	1.97	3.06	4.22	6.09	7.03
2000	0.99	1.39	2.17	2.98	4.30	4.97
3000	0.81	1.14	1.77	2.43	3.51	4.06
5000	0.63	0.88	1.37	1.89	2.72	3.14
8000	0.49	0.70	1.08	1.49	2.15	2.48
10000	0.44	0.62	0.97	1.33	1.92	2.22
13000	0.39	0.55	0.85	1.17	1.69	1.95
15000	0.36	0.51	0.79	1.09	1.57	1.81
17000	0.34	0.48	0.74	1.02	1.48	1.70
22000	0.30	0.42	0.65	0.90	1.30	1.50
26000	0.27	0.39	0.60	0.83	1.19	1.38
30000	0.26	0.36	0.56	0.77	1.11	1.28
50000	0.20	0.28	0.43	0.60	0.86	0.99
80000	0.16	0.22	0.34	0.47	0.68	0.79
100000	0.14	0.20	0.31	0.42	0.61	0.70
130000	0.12	0.17	0.27	0.37	0.53	0.62
150000	0.11	0.16	0.25	0.34	0.50	0.57
180000	0.10	0.15	0.23	0.31	0.45	0.52
200000	0.10	0.14	0.22	0.30	0.43	0.50
230000	0.09	0.13	0.20	0.28	0.40	0.46
240000	0.09	0.13	0.20	0.27	0.39	0.45
250000	0.09	0.12	0.19	0.27	0.38	0.44
256000	0.09	0.12	0.19	0.26	0.38	0.44

[1] To account for sample attrition, multiply the standard error of the estimate by 1.04 for estimates which include data from Wave 5 and beyond.

Table 13. 1993 Topical Module Generalized Variance Parameters[1]

	a	b
Fertility		
# Women	-0.0000437	4425
Births	-0.0000797	8068
Educational Attainment[2]		
Wave 2	-0.0000309	6027
Wave 5	-0.0000337	6569
Wave 8	-0.0000337	6569
Marital Status and Person's Family Characteristics		
Some HH members	-0.0000468	9120
All HH members	-0.0000432	11082
Child Support		
Wave 3	-0.0000723	7319
Support for Non-household Members		
Wave 3	-0.0000375	7319
Health and Disability	-0.0000338	8687
0 - 15 Child Care		
Wave 3	-0.0000959	5922
Welfare History and AFDC		
Both Sexes 18+	-0.0000848	15937
Males 18+	-0.0001767	15937
Females 18+	-0.0001629	15937
Assets and Liabilities		
Wave 4	-0.0000309	6027
Wave 7	-0.0000337	6569

---

[1]   Use the "16+ Income and Labor Force" core parameter for tabulations of reasons for not working/reservation wage and work related income.

[2]   The parameter also applies to the School Enrollment and Finance Topical Module Subject.

Table 14. SIPP 1992, 1993 Combined Panel Topical Module Generalized Variance Parameters

	a	b
Educational Attainment		
1992 wave 8 / 1993 wave 5	-0.0000169	3288
Assets and Liabilities		
1992 wave 7 / 1993 wave 4	-0.0000162	3162
0 - 15 Child Care		
1992 wave 6 / 1993 wave 3	-0.0000503	3107
Child Support		
1992 wave 6 / 1993 wave 3	-0.0000379	3840
Support for Non-household Members		
1992 wave 6 / 1993 wave 3	-0.0000197	3840
Health and Disability		
1992 wave 6 / 1993 wave 3	-0.0000178	4557
1992 wave 7 / 1993 wave 4	-0.0000178	4557

Table 15. Distribution of Monthly Cash Income Among Persons 25 to 34 Years Old

	Total	under $300	$300 to $599	$600 to $899	$900 to $1,199	$1,200 to $1,499	$1,500 to $1,999	$2,000 to $2,499	$2,500 to $2,999	$3,000 to $3,499	$3,500 to $3,999	$4,000 to $4,999	$5,000 to $5,999	$6,000 and over
Thousands in interval	39,851	1371	1651	2259	2734	3452	6278	5799	4730	3723	2519	2619	1223	1493
Percent with at least as much as lower bound of interval	--	100.0	96.6	92.4	86.7	79.9	71.2	55.5	40.9	29.1	19.7	13.4	6.8	3.7

Table 16. SIPP Factors to be Applied to the 1993 Base Parameters to Obtain the 1992, 1993 Combined Panel Parameters

Waves to be Combined

1992 Panel	1993 Panel	g factor[1]
8	5	0.5006
7	4	0.5246
6	3	0.5246

---

[1] When deriving estimates based on two or more waves of data from the same panel, choose the corresponding g-factor with the greatest value. Apply only this factor to the base parameter.

Table 17. Factors to be Applied to Base Parameters to Obtain Combined Panel Parameters for Estimates[1] from Various Reference Periods.

# of available rotation months for 2 panels combined[2]	factor
Monthly Estimate	
2	4.0000
3	3.0000
4	2.0000
5	1.6667
6	1.3333
7	1.1667
8	1.0000
Quarterly Estimates	
12	1.8519
15	1.5631
18	1.2222
19	1.1470
24	1.0000
Annual Estimates	
96	1.0000

[1] Estimates are based on monthly averages.

[2] The number of available rotation months for a given estimate is the sum of the number of rotations available for each month of the estimate for the two panels. There must be at least one rotation month available for each month from each panel for monthly and quarterly estimates.